

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-032422

(43)Date of publication of application : 28.01.2000

(51)Int.Cl.

H04N 7/15  
G06F 3/16

(21)Application number : 10-193145

(71)Applicant : NEC CORP

(22)Date of filing : 08.07.1998

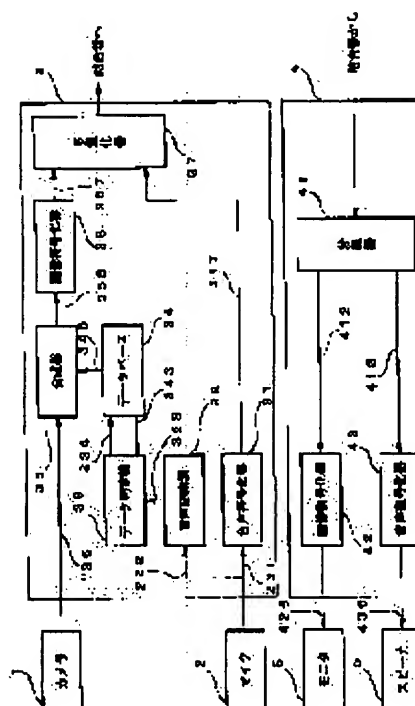
(72)Inventor : KAMURA YUKARI

## (54) IMAGE TRANSMITTER

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide an image transmitter that recognizes voice data to be transmitted selects data corresponding to the voice data from a database, and transmits synthesized data between transmission video data and the voice data.

**SOLUTION:** A voice coder 31 codes a voice signal from a microphone 2, the coded signal is fed to a multiplexer 37, and a voice recognition device 32 recognizes the signal and gives the signal to a data discrimination device 33. The data discrimination device 33 selects the data corresponding to the voice signal from a database 34 and gives the selected data to a synthesizer 35. The synthesizer 35 synthesizes data with a video image from a camera 1 an image coder 36 codes the synthesized data and the coded data are fed to a multiplexer 37. The multiplexer 37 multiplexes an image from the image coder 36 with the voice data from the voice coder 31. A demultiplexer 41 demultiplexes the received multiplexed signal into the coded video image and the coded voice data, the coded video image is decoded by an image decoder 42 and the decoded image is displayed on a monitor 5, and the coded voice data are decoded by a voice decoder 43 and the decoded voice data are outputted to a loudspeaker 6.



## LEGAL STATUS

[Date of request for examination] 08.07.1998

[Date of sending the examiner's decision of rejection] 05.12.2000

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-32422

(P2000-32422A)

(43) 公開日 平成12年1月28日 (2000.1.28)

(51) IntCl. <sup>7</sup>	識別記号	F I	テマコード (参考)
H 0 4 N 7/15		H 0 4 N 7/15	5 C 0 6 4
G 0 6 F 3/16	3 2 0	G 0 6 F 3/16	3 2 0 A
			3 2 0 H

審査請求 有 請求項の数 2 O L (全 6 頁)

(21) 出願番号 特願平10-193145

(22) 出願日 平成10年7月8日 (1998.7.8)

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 加村 ゆかり

東京都港区芝五丁目7番1号 日本電気株式会社内

(74) 代理人 100108578

弁理士 高橋 昭男 (外3名)

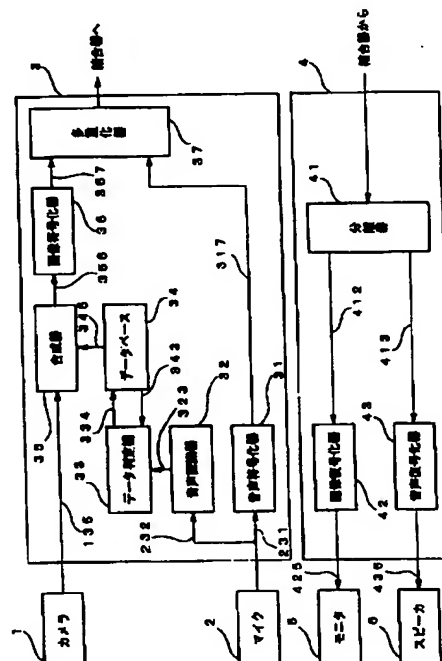
Fターム (参考) 5C064 AAD2 ACO2 ACO6 ACO8 AC17  
AD02 AD09

(54) 【発明の名称】 画像伝送装置

(57) 【要約】

【課題】 送信する音声を音声認識し、その音声に対応したデータをデータベースより選択し、送信映像と合成して送信する画像伝送装置を提供する。

【解決手段】 マイク2からの音声は音声符号化器31で符号化されて多重化器37へ供給されると同時に、音声認識器32で認識されてデータ判定機33へ供給される。データ判定機33によりこの音声に対応するデータがデータベース34より選択され、合成器35へ供給される。合成器35で前記データがカメラ1からの映像と合成され、画像符号化器36で符号化されて多重化器37へ供給される。多重化器37では画像符号化器36からの画像と音声符号化器31からの音声を多重化する。分離器41では、受信された多重化信号が符号化映像と符号化音声に分離され、符号化映像は画像復号化器42で復号化されてモニター5へ、符号化音声は音声復号化器43で復号化されてスピーカ6へ出力される。



## 【特許請求の範囲】

【請求項1】 送信された音声を音声認識する音声認識手段と、  
前記音声認識手段の認識結果に対応するデータを蓄積している蓄積手段と、  
前記データを前記蓄積手段より選択する選択手段と前記選択手段より選択された前記データと前記送信された画像とを合成する合成手段とを具備し、音声に対応したデータを映像と共に送信することを特徴とする画像伝送装置。

【請求項2】 前記蓄積手段には、予め手話映像が蓄積されていることを特徴とする請求項1記載の画像伝送装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、テレビ会議システムにおいて、音声認識を用いてデータの選択を行い、映像と合成して送信する画像伝送装置に関する。

## 【0002】

【従来の技術】一般に、テレビ会議システムを使用する場合、会議中相手の映像を見ていることが多いが、音声だけでは分かりにくいような内容の場合、補足的に映像情報を用いることが会議の運用に有効である。従来のテレビ会議システムにおいては、補足情報として映像情報、例えば図面などをデータとして送って表示させる場合、予め使用するデータを準備しておかななくてはならなかった。

## 【0003】

【発明が解決しようとする課題】しかしながら、従来のテレビ会議システムでは、会議の展開に応じた適切な映像情報をタイムリーに使用するのには難しいという問題があった。その理由は、送信するデータは予め用意されたものか、送信されている映像や音声の情報に左右されない内容のものを映像や音声と多重化して送信することし

かできないからである。  
【0004】本発明はこのような点を考慮してなされたもので、テレビ会議において、送信する音声

## 【0005】

【課題を解決するための手段】請求項1記載の発明は、テレビ会議システムにおいて、送信された音声

## 【0006】

【発明の実施の形態】以下、図面を参照して本発明の一実施形態について説明する。図1は本発明の一実施形態による画像伝送装置の構成を示すブロック図である。この図において、3はCODEC送信部、4はCODEC受信部である。CODEC送信部3において、符号31はマイク2から供給された音声を符号化する音声符号化器、32はマイク2から供給された音声を認識する音声認識器、33は音声認識器32で認識された音声に対応するデータをデータベース34より選択するデータ判定機、35はカメラ1より供給された映像とデータベース34より供給されたデータとを合成する合成器、36は合成器35より供給された映像を符号化する画像符号化器、37は音声符号化器31から供給された符号化音声を画像符号化器36から供給された符号化画像と多重化する多重化器である。また、CODEC受信部4において、41は符号化音声と符号化画像の多重化信号を符号化映像と符号化音声とに分離する分離器、42は分離器41からの符号化画像を復号化して線425を介してモニタ5へと出力する画像復号化器、そして43は分離器41からの符号化音声を復号化して線436を介してスピーカ6へと出力する音声復号化器である。

【0007】次に、上記構成による画像伝送装置の動作について説明する。カメラ1からの映像は、線135を介して合成器35に供給される。マイク2からの音声は、線231を介して音声符号化器31へ、また線232を介して音声認識器32へと供給される。音声符号化器31では供給された音声を符号化し、線317を介して多重化器37へと供給する。音声認識器32では供給された音声を認識し、そのデータを線323を介してデータ判定機33へと供給する。データ判定機33では認識された音声に対応するデータを線343を介してデータベース34より選択し、その判定結果を線334を介して再びデータベース34へ供給する。データベース34は、データ判定器33から供給された情報に対応するデータを線345を介して合成器35へ供給する。合成器35ではカメラ1より供給された映像とデータベース34より供給されたデータとを合成し、線356を介して画像符号化器36へと供給する。画像符号化器36では供給された映像を符号化し、線367を介して多重化器37へと供給する。多重化器37では画像符号化器36から供給された符号化画像と音声符号化器31から供給された符号化音声を多重化し、結合器へ出力する。

【0008】分離器41では結合器を介して供給された多重化信号を符号化映像と符号化音声とに分離し、符号化映像は線412を介して画像復号化器42へ供給され、符号化音声は線413を介して音声復号化器43へと供給される。画像復号化器42では符号化映像を復号化して線425を介してモニタ5へと供給する。モニタ5では送信された映像を表示する。音声復号化器43では符号化音声を復号化して線436を介してスピーカ6へと供給す

る。

【0009】図2は、複数の画像伝送装置によって構成されるテレビ会議システムの構成を示す図である。複数の装置間におけるデータの送受信は、ISDN(Integrated Service Digital Network)などのネットワークを介して行われる。装置A側のカメラ1からの映像およびマイク2からの音声は、CODEC送信部3に供給される。これらの映像および音声は、CODEC送信部3において前述のような処理を施されて多重化されたデータとして結合器7へ供給される。結合器7へ供給されたデータは、ISDN8を介して装置B側の結合器7'へ供給される。

【0010】結合器7'に供給されたデータは、装置B側のCODEC受信部4'へ供給され、そこで前述のように、再び映像データと音声データとに分離され、さらに復号化されて、モニタ5'およびスピーカ6'へそれぞれ供給される。以上は装置A側から装置B側へのデータの流れについて述べたが、装置B側から装置A側へのデータの流れについても全く同様であるので、これについての説明は省略する。

【0011】本発明により、データベースに手話の映像データを蓄積しておき、音声認識器から送信された音声データに対応した手話の映像データをデータ判定器がデータベースより選択することによって、耳の不自由な人に手話を知らない人の会話を聞かせることが可能になる。また、データベースに他国語のデータを蓄積しておき、送信する言葉に対応する翻訳語をデータ判定器が選択することによって、異なる言語を話す人同士の会話が可能となる。

【0012】

\*【発明の効果】以上説明したように、本発明によれば、送信される音声を音声認識手段により認識し、その音声に対応したデータを蓄積手段より選択し、送信された映像と合成して送信するので、送信側の音声に対応したデータを、音声と同時に受信側に表示できるという効果が得られる。

【図面の簡単な説明】

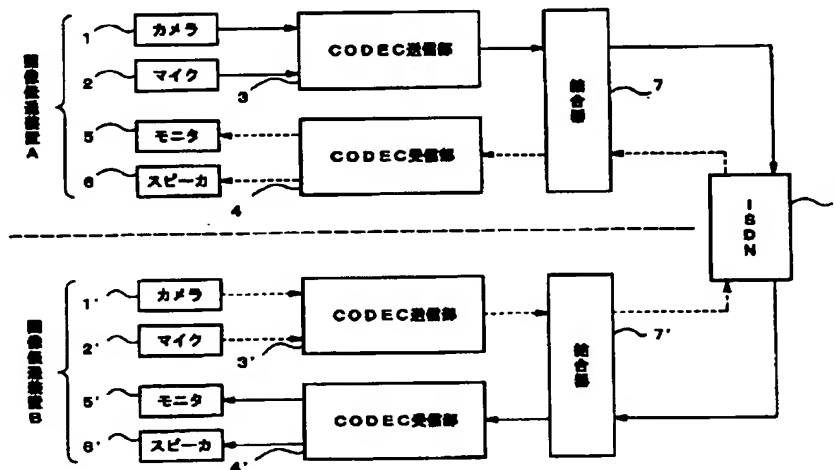
【図1】 本発明の一実施形態による画像伝送装置の構成を示すブロック図である。

【図2】 本発明の一実施形態による、複数の画像伝送装置間でのデータのやりとりを示すブロック図である。

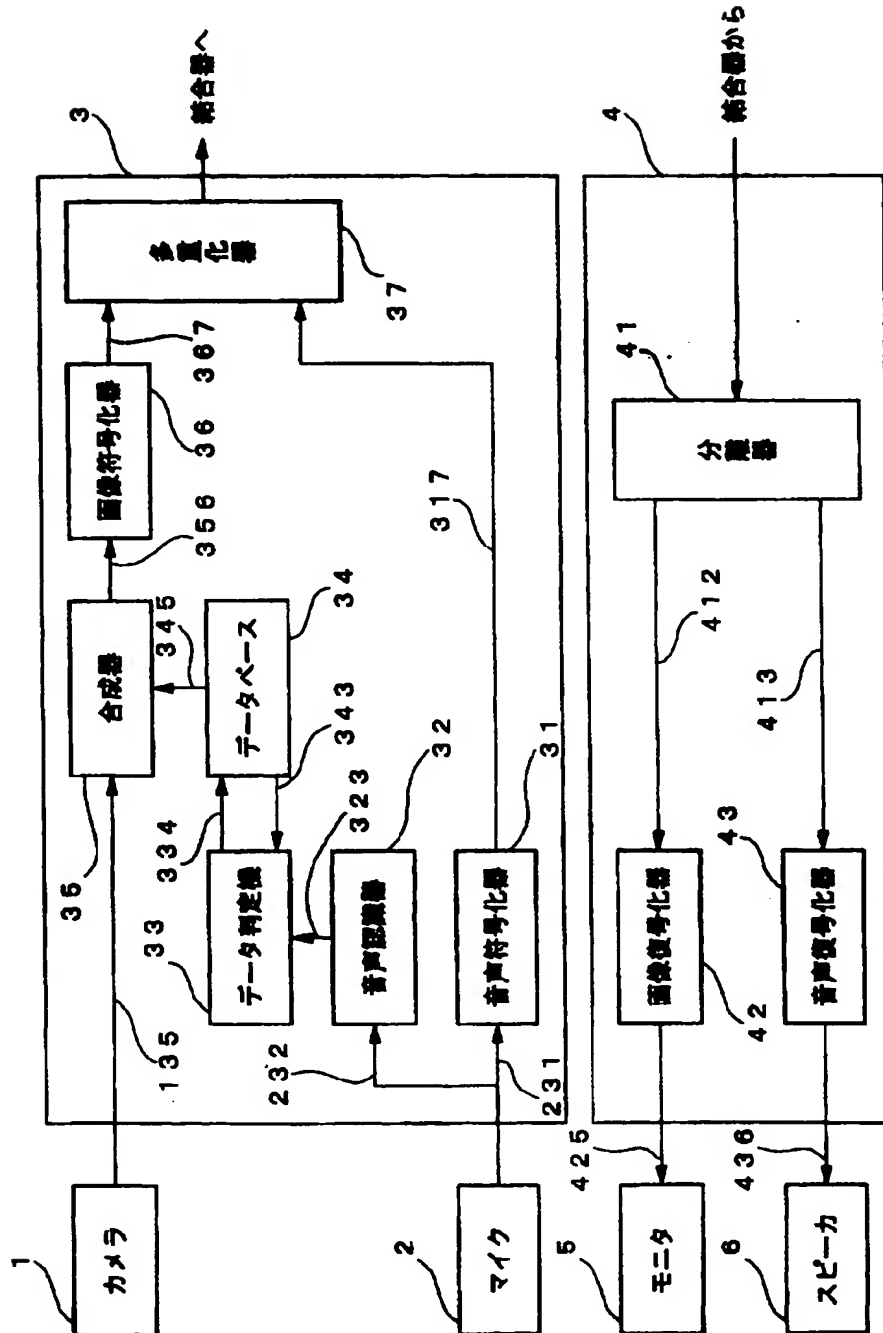
【符号の説明】

- 1, 1' . カメラ
- 2, 2' . マイク
- 3, 3' . CODEC送信部
- 31. 音声符号化器
- 32. 音声認識器
- 33. データ判定機
- 34. データベース
- 35. 合成器
- 36. 画像符号化器
- 37. 多重化器
- 4, 4' . CODEC受信部
- 41. 分離器
- 42. 画像復号化器
- 43. 音声復号化器
- 5, 5' . モニタ
- 6, 6' . スピーカ
- 7, 7' . 結合器
- \*30 8. ISDN

【図2】



【図1】



【手続補正書】

【提出日】平成10年8月19日(1998. 8. 1

\*【補正対象項目名】全図

9)

【手続補正1】

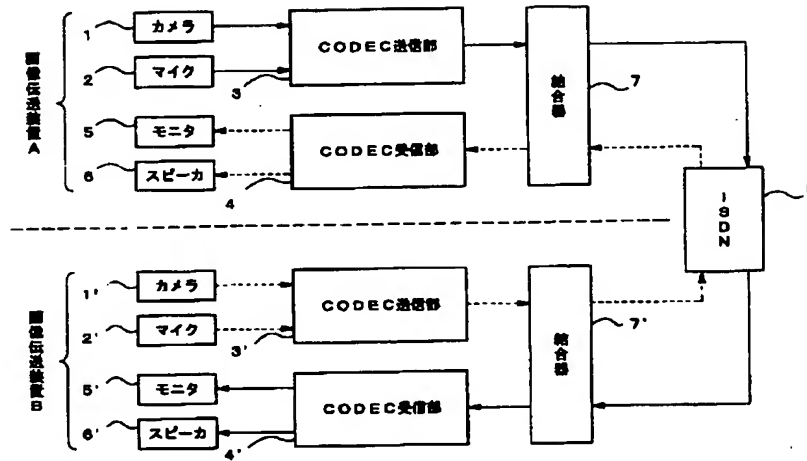
【補正方法】変更

【補正対象書類名】図面

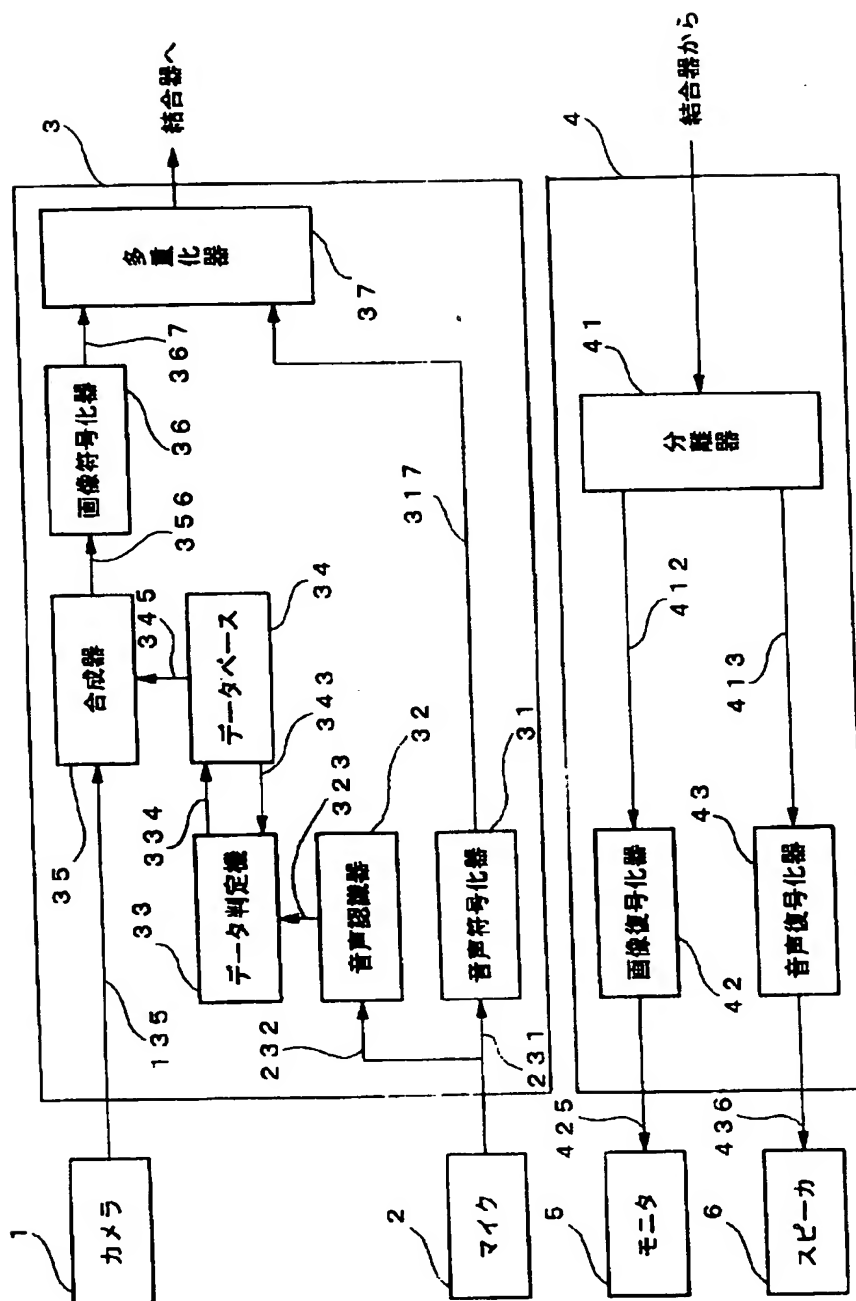
\*

【補正内容】

【図2】



【圖 1】





\* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

CLAIMS

---

[Claim(s)]

[Claim 1] The picture transmission equipment characterized by to provide a synthetic means compound said data chosen from the speech-recognition means which carries out [ voice / which was transmitted ] speech recognition, an are-recording means to by\_which the data corresponding to the recognition result of said speech-recognition means store, a selection means choose said data from said are-recording means, and said selection means, and said image which were transmitted, and to transmit the data corresponding to voice with an image.

[Claim 2] Picture transmission equipment according to claim 1 characterized by accumulating the sign language image in said are recording means beforehand.

---

[Translation done.]

\* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] In a video conference system, this invention

chooses data using speech recognition, and relates to the picture transmission equipment which compounds with an image and is transmitted.  
[0002]

[Description of the Prior Art] Although a partner's image is generally seen during the meeting in many cases when using a video conference system, in the case of contents which are unclear, it is effective in employment of a meeting only with voice to use image information additionally. In the conventional video conference system, when image information, for example, a drawing etc., was sent as data and it was displayed as extra information, the data used beforehand had to be prepared.

[0003]

[Problem(s) to be Solved by the Invention] However, in the conventional video conference system, there was a problem that it was difficult to use timely the suitable image information according to expansion of a meeting. It is because multiplexing what was prepared beforehand, and the thing of the contents which are not influenced by the information on the image transmitted or voice with an image and voice, and transmitting can only do the data which transmit the reason.

[0004] This invention was made in consideration of such a point, recognizes the voice which transmits by speech recognition at a television conference, and chooses the data corresponding to the voice from a database, and it aims at offering the picture transmission equipment which compounds with a transmitting image and is transmitted.

[0005]

[Means for Solving the Problem] A speech recognition means by which invention according to claim 1 carries out [ voice / which was transmitted ] speech recognition in a video conference system, An are recording means by which the data corresponding to the recognition result of said speech recognition means are stored, It is picture transmission equipment characterized by providing a synthetic means to compound the data chosen from a selection means to choose said data from said are recording means, and said selection means, and the transmitted image, and transmitting the data corresponding to voice with an image.

[0006]

[Embodiment of the Invention] Hereafter, 1 operation gestalt of this invention is explained with reference to a drawing. Drawing 1 is the block diagram showing the configuration of the image transmission equipment by 1 operation gestalt of this invention. In this drawing, 3 is the CODEC transmitting section and 4 is a CODEC receive section. The voice encoder which encodes the voice to which the sign 31 was supplied

from the microphone 2 in the CODEC transmitting section 3, The speech recognition machine which recognizes the voice to which 32 was supplied from the microphone 2, the data judging machine which chooses from a database 34 the data corresponding to the voice 33 has been recognized to be with the speech recognition vessel 32, The synthetic vessel which compounds the image to which 35 was supplied from the camera 1, and the data supplied from the database 34, The image encoder which encodes the image to which 36 was supplied from the synthetic vessel 35, and 37 are the coded image to which the coding voice supplied from the voice encoder 31 was supplied from the image encoder 36, and a multiplexing machine to multiplex. Moreover, in the CODEC receive section 4, they are the eliminator with which 41 divides the multiplexed signal of coding voice and a coded image into a coding image and coding voice, the image decryption machine which 42 decrypts the coded image from an eliminator 41, and is outputted to a monitor 5 through a line 425, and the voice decryption machine which 43 decrypts the coding voice from an eliminator 41, and is outputted to a loudspeaker 6 through a line 436.

[0007] Next, actuation of the picture transmission equipment by the above-mentioned configuration is explained. The image from a camera 1 is supplied to the synthetic vessel 35 through a line 135. It passes through the voice from a microphone 2 voice encoder 31 through a line 231, and it is supplied to the speech recognition machine 32 through a line 232. In the voice encoder 31, the supplied voice is encoded and the multiplexing machine 37 is supplied through a line 317. With the speech recognition vessel 32, the supplied voice is recognized and the data is supplied to the data judging machine 33 through a line 323. In the data judging machine 33, the data corresponding to the recognized voice are chosen from a database 34 through a line 343, and the judgment result is again supplied to a database 34 through a line 334. A database 34 supplies the data corresponding to the information supplied from the data judging machine 33 to the synthetic vessel 35 through a line 345. With the synthetic vessel 35, the image supplied from the camera 1 and the data supplied from the database 34 are compounded, and the image encoder 36 is supplied through a line 356. In the image encoder 36, the supplied image is encoded and the multiplexing machine 37 is supplied through a line 367. With the multiplexing vessel 37, the coding voice supplied from the coded image supplied from the image encoder 36 and the voice encoder 31 is multiplexed, and it outputs to a coupler.

[0008] In an eliminator 41, the multiplexed signal supplied through the coupler is divided into a coding image and coding voice, a coding image is supplied to the image decryption machine 42 through a line 412, and

coding voice is supplied to the voice decryption machine 43 through a line 413. With the image decryption vessel 42, a coding image is decrypted and a monitor 5 is supplied through a line 425. The transmitted image is displayed in a monitor 5. With the voice decryption vessel 43, coding voice is decrypted and a loudspeaker 6 is supplied through a line 436.

[0009] Drawing 2 is drawing showing the configuration of the video conference system constituted by two or more image transmission equipment. Transmission and reception of the data between two or more equipments are performed through networks, such as ISDN (IntegratedService Digital Network). The voice from the image and microphone 2 from a camera 1 by the side of Equipment A is supplied to the CODEC transmitting section 3. Such images and voice are supplied to a coupler 7 as data which the above processings were performed and were multiplexed in the CODEC transmitting section 3. The data supplied to the coupler 7 are supplied to coupler 7' by the side of Equipment B through ISDN8.

[0010] The data supplied to coupler 7' are supplied to CODEC receive section 4' by the side of Equipment B, as mentioned above, it separates into image data and voice data again, and are decrypted further, and are supplied to monitor 5' and loudspeaker 6' there, respectively. Although the above described the data flow from Equipment A side to Equipment B side, since the same is completely said of the data flow from Equipment B side to Equipment A side, the explanation about this is omitted.

[0011] When the image data of sign language are stored in the database by this invention and a data judging machine chooses from a database the image data of the sign language corresponding to the voice data transmitted from the speech recognition machine, it becomes possible to tell a person hard of hearing the conversation of those who do not know sign language. Moreover, when the data of foreign country language are stored in the database and a data judging machine chooses the translation word corresponding to the language to transmit, the conversation of those which speak different language is attained.

[0012]

[Effect of the Invention] Since according to this invention recognize the voice transmitted with a speech recognition means, the data corresponding to the voice are chosen from an are recording means, it compounds with the transmitted image and it transmits as explained above, the effectiveness that the data corresponding to the voice of a transmitting side can be displayed on voice and coincidence at a receiving side is acquired.

---

[Translation done.]

\* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

DESCRIPTION OF DRAWINGS

---

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram showing the configuration of the picture transmission equipment by 1 operation gestalt of this invention.

[Drawing 2] It is the block diagram which is twisted in 1 operation gestalt of this invention and in which showing an exchange of the data between two or more picture transmission equipments.

[Description of Notations]

- 1 and 1'. camera
  - 2 and 2'. microphone
  - 3 and 3'. CODEC transmitting section
  - 31. Voice Encoder
  - 32. Speech Recognition Machine
  - 33. Data Judging Machine
  - 34. Database
  - 35. Synthetic Vessel
  - 36. Image Encoder
  - 37. Multiplexing Machine
  - 4 and 4'. CODEC receive section
  - 41. Eliminator
  - 42. Image Decryption Machine
  - 43. Voice Decryption Machine
  - 5 and 5'. monitor
  - 6 and 6'. loudspeaker
  - 7 and 7'. coupler
  - 8. ISDN
-

[Translation done.]

\* NOTICES \*

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

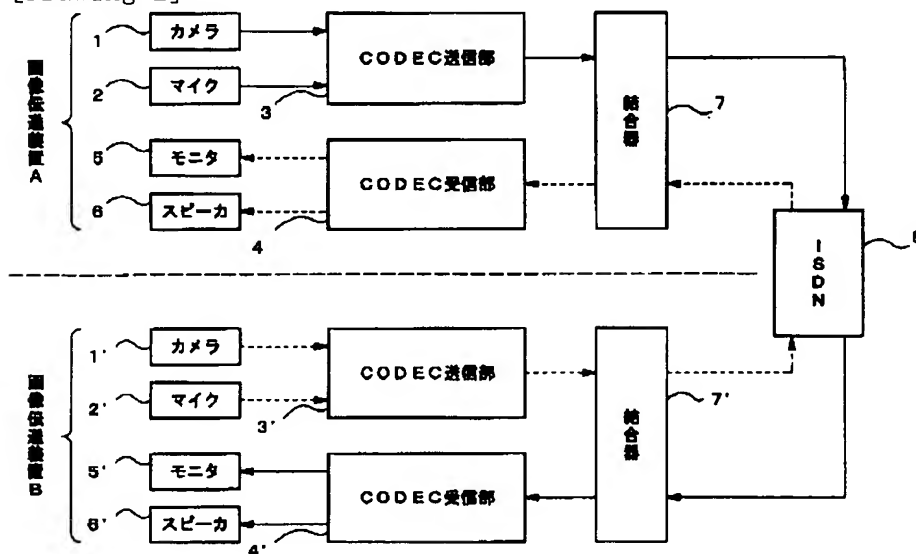
1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

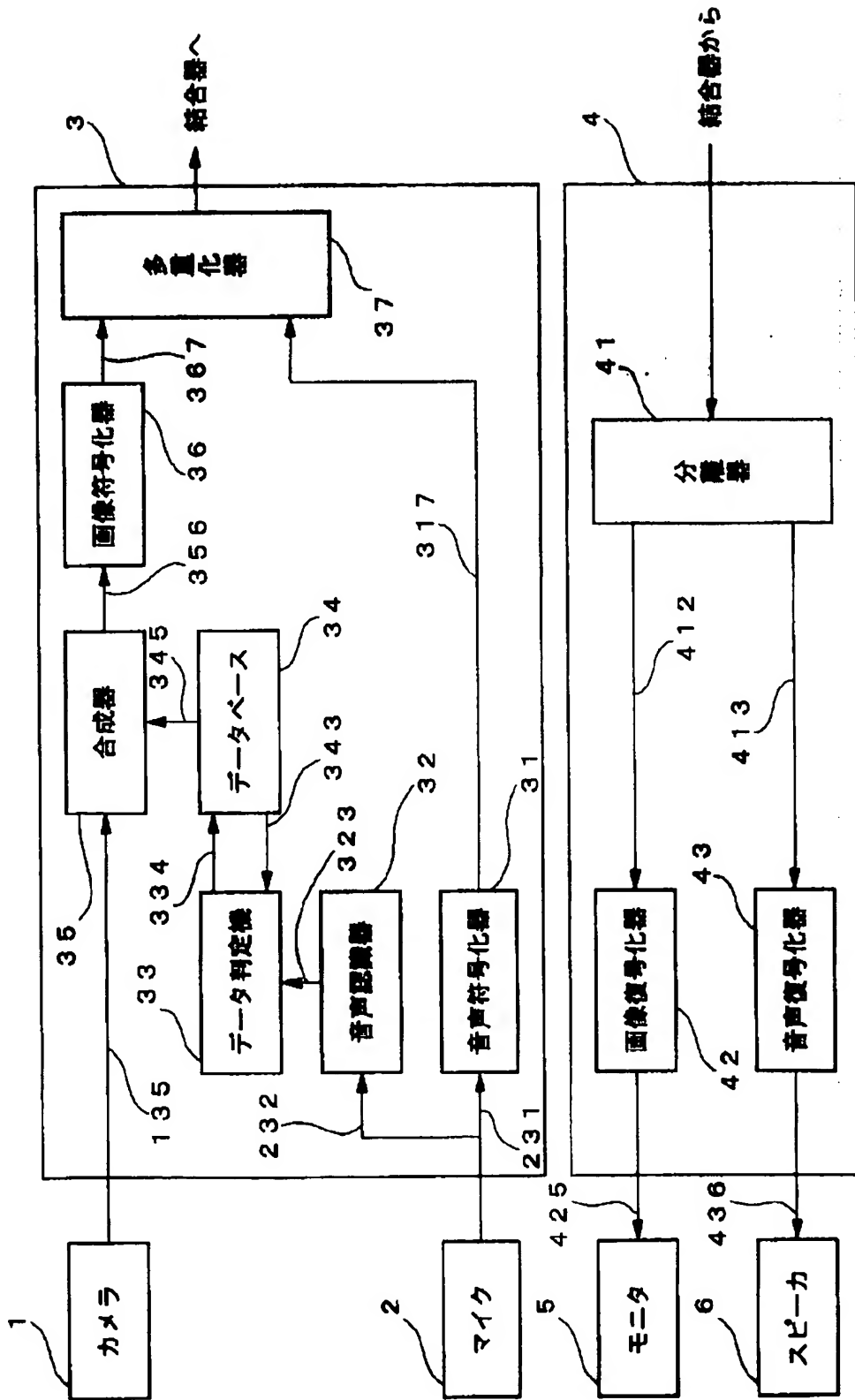
## DRAWINGS

---

[Drawing 2]



[Drawing 1]



[Translation done.]